DEVELOPMENT OF A HYBRID ANNOTATION SYSTEM FOR EXTRACTING VISUAL INFORMATION FROM MULTIMODAL DATA

Synopsis

Submitted in the partial fulfilment for the requirement of

Doctor of Philosophy

In

Physics and Computer Science

Submitted by

Ms. Manisha Gupta

Under the supervision of

Prof. Preetvanti Singh

DEPARTMENT OF PHYSICS AND COMPUTER SCIENCE, FACULTY OF SCIENCE

DAYALBAGH EDUCATIONAL INSTITUTE (DEEMED TO BE UNIVERSITY)

DAYALBAGH, AGRA (282005)

2022

INTRODUCTION

The web, today, has tremendous growth in various formats of data like video, PDFs, images, and text. Extracting meaningful information from this data is almost impossible. Thus, there is a need for a technology that provides an understanding for the visual environment to the computers. Computer vision, a branch of artificial intelligence, assists in automating visual understanding.

Computer Vision

Computer Vision, by utilizing machine learning tools, assists in developing an algorithmic basis to achieve automatic visual understanding (Ballard, 1989; Trucco & Verri, 1998). Computer vision can identify, track, measure, detect, and classify objects in an image or video, which can assist in making useful decisions for the visual environment. Applications of this approach are found in facial recognition, self-driving cars, construction safety assurance and extracting data from images and PDFs.

Computer Vision Techniques

In a way, Computer vision is all about pattern recognition. The techniques that allow the computer to hunt down patterns are:

1. **Image classification**

   Image classification is the process of predicting a specific class or label for digital images that are defined by a set of data points. It is a subset of the classification problem, where an entire image is assigned a label. Convolutional neural networks are a method for image classification.

2. **Object detection**

   Object detection is used to define objects within images that involve outputting bounding boxes and labels for individual objects. This is done by applying classification and localization to many objects instead of just a single dominant object.

3. **Object Tracking**

   Object tracking refers to the process of following an object(s) of interest in a given scene. The methods for object tracking can be done using a generative model to describe the apparent characteristics or by using the discriminative method to distinguish between the objects.

4. **Semantic Segmentation**

   Semantic segmentation is used to semantically understand the role of each pixel in the image. It is a type

of image annotation that helps in associating labels or keywords based on a cluster of pixels. Image annotation is an area which is used to understand the semantic segmentation algorithms.

5. **Instance Segmentation**

This technique segments different instances of classes, such as labeling 5 cars with 5 different colors and shapes. Image annotation is a sub-system of computer vision that assists the above tasks to assign meaningful keywords to the objects present in a scene automatically.

Image Annotation

With the rapid development of the Internet and mobile technologies, the number of digital images has grown exponentially. Efficiently classifying the objects in such a huge image database is very difficult, and may lead to inefficient searching and indexing of the images. For example, searching an image having a cat in an unorganized large image database becomes inefficient as it takes a long response time. Also due to imprecise and superficial annotation, the output may be irrelevant images with an animal such as tiger, lion, fox, wolf. Image annotation plays a vital role in dealing with labeling the images in the exponentially growing digital images.

Image annotation is an important task in artificial intelligence that gives machine eyes, i.e. the ability to see and interpret the world. It is a process of labeling or classifying the images with text and/or annotation tools to show the data features you want your model to recognize on its own. Annotated images are labeled images based on the semantics of objects, in which every object is annotated with the correct representation of it. Users can easily recognize critical semantics based on the keywords given for an image. Figure 1 shows an annotated image (Hanbury, 2006), Figure 1(a) is an example of whole image annotation where the keywords are associated with the image; whereas Figure 1(b) is an example of segmentation and annotation, where keywords are associated with each region of the segmentation. Image annotation technique is of great interest for the researchers as it helps in

- efficient retrieval of image content from a largescale image database,
- classification of different objects,
- auto-illustration of the image, and
- automatic arrangement of labels.

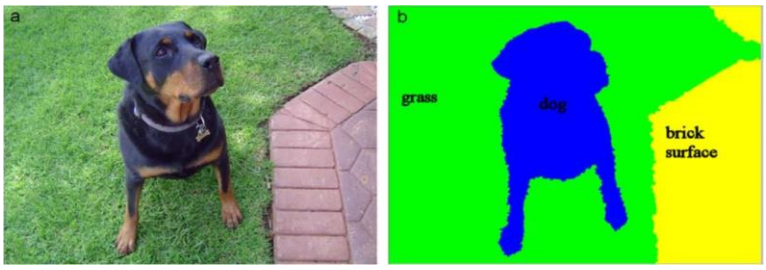

Figure 1: Image annotation

Image annotation can be done in three ways:

- **Manual**: Manual annotation includes entering descriptive keywords by the users when the images are loaded, registered, or browsed. This method is considered best in terms of accuracy as keywords for the images are selected based on human determination for their semantic content. The main disadvantage of the method is that if the users are retrieving the images after a long period, they may forget the annotations they used.
- **Semi-Automatic Annotation:** Semi-automatic annotation requires some sort of user participation in the image annotation process. It explains machine learning algorithms for user-supported image annotation. It is a consistent, cost-effective, fast, and intelligent method for annotation of visual data. This method is particularly suitable in a dynamic database system, in which new images are continuously being introduced (Jin, Li, Ma, Guo & Yu, 2017; Schleier et al., 2020).
- **Automatic Image Annotation:** Automatic image annotation allows the system to automatically assign metadata in the form of keywords to a digital image. The modeling framework of automatic image annotation is based on nonparametric density estimation using the technique of kernel smoothing which results in more accurate annotations. This approach is fastest also among all. The automatic image annotation system utilizes existing annotated image datasets to tie visual features with keywords using machine learning techniques and to predict the keyword for any given un-annotated images. It is used in image retrieval systems to organize and locate images of interest from the image database (Hidajat, 2015; Vatani, Ahvanooey & Rahimi, 2020).

The Image Annotation Framework

The image annotation framework is shown in Figure 2. Raw images (i.e. collection of images in the database that are not correctly indexed), when processed through image annotation tools, result in annotated images that provide meaningful information that can help in efficient decision making. The

image annotation tools are used to label image contents in a way that the result is closer to appropriate query representation. The labeling is done based on:

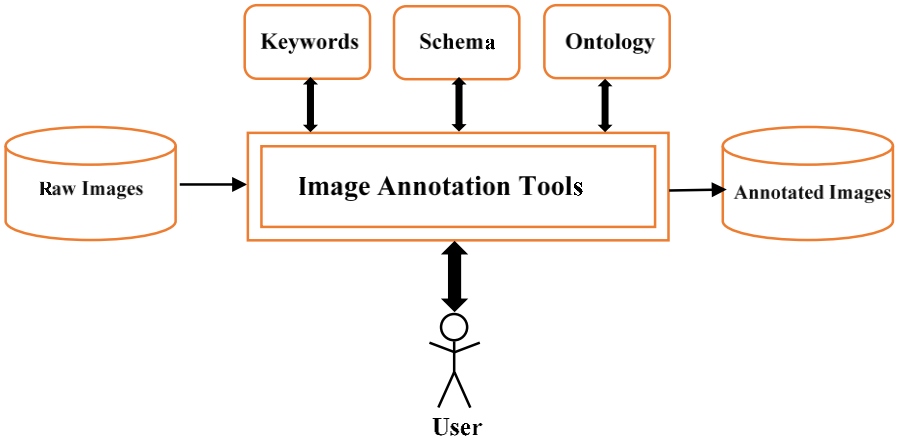

Figure 2: Image Annotation Framework

- **Keywords:** Keywords are the labels that uniquely identify all the objects of the image based on its semantics and distinguishable attributes.
- **Schema:** Schema is defined as the logical design and structure of the model.
- **Ontology:** Ontology is a design of an abstract that defines a collection of concepts. The focus of ontology-based semantic image annotation is on relating the contents of an image and describing image contents as completely as possible.

Types of image annotation

To label image datasets, image annotations can be of following types:

Table 1: Image Annotation Types

S.No.	Type of Image Annotation	Description
1.	**Bounding Boxes**	Most used and simplest. Requires labelers to draw a box as close as possible to the edges of key objects within the image. Used for object classification, localization, and detection in industries such as retail, e-commerce, and healthcare.
2.	**3D Cuboids**	Depicts label length, width, and approximate depth. Requires drawing a box encapsulating the object of interest and placing anchor points at each of the object's edges.
3.	**Polygons**	Used for more precise annotation of items that are irregularly shaped. Requires a high level of precision from the labeler.

S.No.	Type of Image Annotation	Description
4.	**Lines and Splines**	• Used to draw boundaries in a region of an image. • Mainly used when a section that needs to be delineated is too small or thin.
5.	**Point Annotation**	• Involves the accurate plotting of key points at a specified location on an image. • Mostly used for facial recognition and sentiment analysis.
6.	**Semantic Segmentation**	• Is the process of associating every single pixel in an entire image with a tag. • Commonly used for medical imaging devices.

Image Annotation Methods

The following are the methods of image annotation used for a different form of an image (Figure 3).

1. **Model-based annotation approach**

The model-based annotation approach trains the annotation model from the training set features (either visual, textual, or both) to annotate the unknown images. In this approach, an annotation model is learned from the training data through visual and textual correlation which then can accurately annotate the unknown image. Visual and textual correlation is a relationship established through visual features and textual metadata (labels and surrounding texts) between a pair of instances or multiple instances. Authors (Tewari et al., 2017; Mehmood, Mahmood & Javid, 2018) used different methods for this approach like conventional neural network and support vector machine.

2. **Learning-based annotation approach**

The visual features are not sufficient to annotate all the objects present in the image as these do not represent the contents of the image in its entirety. In order to label the contents of the image accurately, the model should learn features from various available sources. Machine learning tools are used for developing learning base annotation (Zhang, Hsu, Lai & Zheng, 2018; Sturman et al., 2020).

3. **Image annotation based on tag length**

This method focuses on the length of the tags associated with the images. These tags may be fixed or variable lengths. When an image is annotated, it is assigned labels based on the contents of the image either using a fixed-length tag or an arbitrary length tag. Image tagger and dataturks tools provide many data annotation capabilities, including image segmentation, named entity recognition tagging in

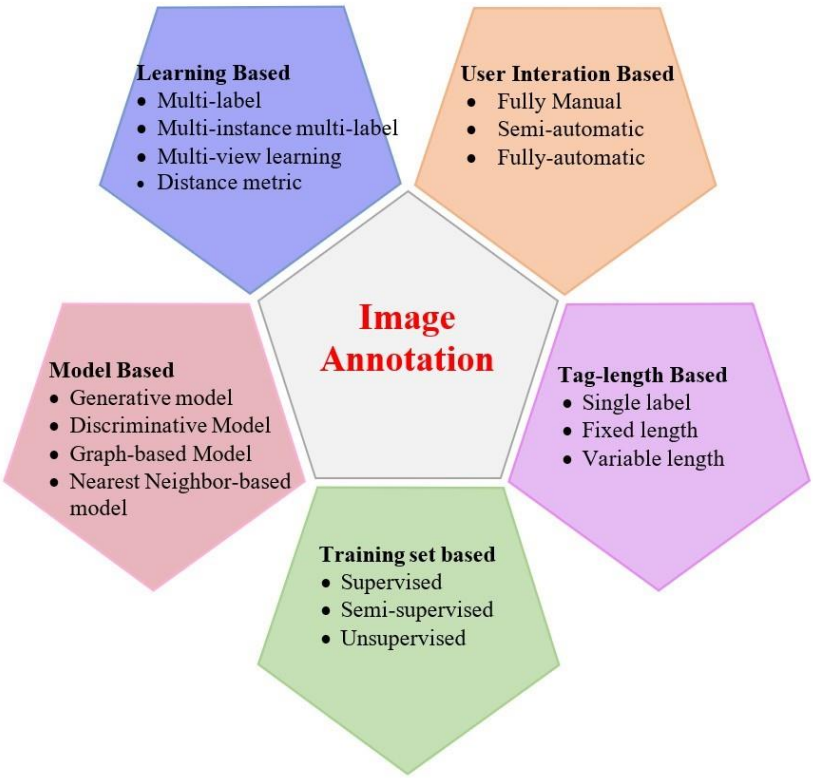

Figure 3: Image annotation methods

documents, and part of speech tagging (Zheng, Caiming & Caixian, 2018; Alqahtani & Alsulaiman, 2020).

4. **Image annotation based on the training dataset**

Consideration of a training data set to design an annotation model is important. This has a significant influence on the final annotation model. The dataset can be categorized into three sets- the training set having complete labels, the training set having incomplete labels, and the training dataset which is not annotated. Various methods are used for the training datasets by Huang, Yang, Huang, Wu & Lin, 2020; Kromp et al., 2020.

5. **Image annotation based on user interaction**

One way to annotate an image is based on the interaction of a user with the system. If the system is fully automatic then there is no interaction with the user. If the system is fully manual, then a human expert performs the annotation which itself is a tedious, time consuming, and very expensive process. Annotation can also be performed in a semi-automatic manner where most of the processing is handled by the system and the users can interact with the system using relevance feedback to improve the

confidence of the model. Pellegrin, Escalante, Gómez & González (2017) and Sbaiz & Vetterli (2018) proposed various approaches for image annotation based on user interaction.

Image Annotation Process

The process of image annotation is shown in Figure 4. First, the metadata is acquired from the image, which includes fields like time, GPS, and artist. This information is then used to analyze valuable information in personal utilities such as email, and the produced results are presented as annotation implications. Based on the derived annotation implication and manually added tags (say emotion) images are annotated. This photo and the annotations are uploaded to the server, which finally accumulates the photo and the annotations and generates multidimensional indices for the photos. Now the server obtains the GPS and other metadata and annotates the upload time and the owner of the photo and creates thumbnails for the photo. Along with this, it also generates multidimensional indices for the photo. These indices are used to retrieve the image.

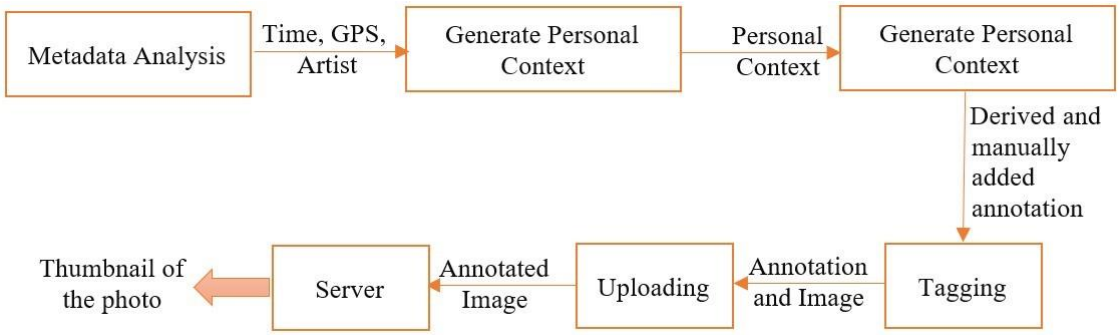

Figure 4: Image Annotation Process

Application of image annotation is found mainly in healthcare (Song, Zhao, Luo & Dou (2017); Yin & Bi (2018); Li et al. (2020); and Waheed et al. (2020)); and environmental monitoring (James, Marzloff, Barrett, Friedman & Johnson (2017) and Zurowietz, Langenkämper, Hosking, Ruhl & Nattkemper (2018)).

REVIEW OF LITERATURE

Researchers have used various annotation methods to provide efficiency in labeling the images. Each of the image annotation technique is reviewed below:

1. **Model-based annotation approach**

This approach trains the annotation model from the training set features. Amiri & Jamzad (2015)

designed a semi-supervised learning framework that, by incorporating unlabeled images into the training phase, reduces the system demand to labeled images. The framework is based on gamma distribution and a modified expectation-maximization algorithm to update parameters of the constructed generative models. Osaim & Bouguila (2015) developed a hybrid generative discriminative framework based on a support vector machine and Gamma mixture. 3D shape class recognition, object categorization, and video event analysis were performed for evaluating the effectiveness of the proposed framework. Jin, Luo & Guo (2015) proposed a feature selection approach to improve the performance of automatic image annotation by introducing a nonlinear factor for its evaluation. Murthy, Maji & Manmatha (2015) developed a model by using a convolutional neural network to extract features from an image and word embedding vectors to represent their associated tags. Ding et al. (2016) developed a context-aware multi-instance, multi-label model to integrate the instance context and label context into a general framework. Ivasic-Kos, Pobar, & Ribaric (2016) dealt with the two-tier image annotation model-based on a multi-label classifier and fuzzy-knowledge representation scheme. Song et al. (2016) proposed a semi-supervised annotation approach by learning an optimized graph from multi-cues to accurately embed the relationships among the data points.

Ji, Gao & Hu (2017) proposed an image annotation approach by combining the generative and discriminative models through local discriminant topics in the neighborhood of the unlabeled image. Singular value decomposition(SVD) was applied to group the images of the neighborhood into different topics according to their semantic labels. Jiu & Sahbi (2017) presented four different frameworks to determine the weights of supervised, unsupervised, kernel-based semi-supervised, and Laplacian-based semi-supervised networks. Li, Shen, Liu & Zhang (2017) developed a rank-preserving low-rank factorization model for image annotation with missing labels to deal with noisy data. Feng & Lang (2018) developed graph regularized low-rank feature mapping for multi-label learning with application to image annotation. The proposed method concatenated the prediction models for different tags into a matrix. Zheng, Caiming & Caixian (2018) implemented a multi-modal data fusion-based latent dirichlet allocation topic model to annotate social images via fusing visual content, user-supplied tags, user comments, and geographic information. Liu et al. (2018) designed a support vector machine-

based multi-label learning system with missing labels for image annotation. Tian (2018) elaborated on the basic principles of the

probabilistic latent semantic analysis model followed by its applications to image annotation, image retrieval, and image classification.

Ke, Zou & Niu (2019) presented an end-to-end automatic image annotation model-based on the deep convolutional neural network that transforms the image annotation problem into a multi-label learning problem. Ma, Xie, Luo & Xiong (2019) developed a multi-label linear discriminant approach for computing weights to improve the accuracy in distance calculation. Song et al. (2020) proposed an annotation method, the local learning-based probabilistic latent semantic analysis, to overcome the shortcoming of the annotation model. Wang, Mueller, Bernard, Theobalt (2020) utilized a model-based generative loss for training hand pose estimators on depth images based on a volumetric hand model.

2. Learning-based annotation approach

This approach for image annotation is used to learn a model from multiple sources of the training set Luo, Liu, Tao & Xu (2015) developed a multiview matrix completion method for the multi-label image classification and transudative (semi-supervised) multilabel classification. Pellegrin, Vanegas, Ovalle & Beltran (2015) presented the concept of semantic image annotation based on the multimodal representation that is built in an unsupervised way by using the associated text of images and the visual features that represent them. Wang et al. (2015) presented weakly semi-supervised deep learning method for multi-label image annotation that enables to train the deep convolutional neural network with images from social networks. Wu, Lyu, Hu & Ji (2015) developed a method for multi-label learning that explicitly handles missing labels. Jing, Wu, Li & Hu (2016) presented a multi-label learning approach with label consistency regularization and partial-identical label embedding. Shi, Yang & Hospedales (2016) designed a model for weakly-supervised image annotation and segmentation using a non-parametric Bayesian model.

Maihami & Yaghmaee (2017) reviewed the application of structured sparse representation for image annotation. Xia, Chen, Li, Zhang & Whang (2017) developed a method for automatic image annotation based on multi-feature fusion and multi-label learning algorithms. Uricchio, Ballan, Seidenari, Bimbo (2017) proposed an image annotation method using label transfer in semantic space,

based on kernel canonical correlation analysis, the correlation of visual and textual features are well preserved into a semantic embedding. Verma & Jawahar (2017) addressed the issue of image annotation by propagating labels from semantic neighborhoods. Wang & Li (2017) designed a multimodal learning framework for automatic image annotation. Guo, Han, Su & Sun (2018) presented a method to train a deep convolutional neural network with images from social networks where images are well labeled, with several labels or uncorrelated labels. Li et al. (2018) presented a multi-modal feature fusion-based framework to improve the geographic image annotation. To achieve effective representations of geographic images, the method leverages a low-to-high learning flow for both the deep and shallow modality features.

Adikhresna, Kusumaningrum, Warsito (2019) presented an experimental study by comparing the performance of various multi-label classification methods for multi learning and multi-label radial basis function. Jin & Jin (2019) developed a multi-label automatic image annotation approach based on machine learning. A distance metric learning method based on cost-sensitive learning for this task was developed to reduce the impact of class imbalance of samples. Srivastava & Srivastava (2020) examined the efficiency framework for image annotation involving salient object detection, feature extraction, feature selection, and multi-label classification. Yakovlev & Lisovychenko (2020) developed a method for image annotation automatization using artificial intelligence models learning.

3. Image annotation based on tag length

Researchers have enhanced the image annotation methods by considering the length of the tags. Im & Park (2015) presented an annotation method that connected image tags using predicate words to capture the contexts in which the image tags are used. Jin & Nakayam (2016) developed a recurrent image annotator model that forms an image annotation task as a sequence generation problem to predict the proper length of tags according to image contents. Li, Uricchio, Ballan & Bertini (2016) discussed the key functionalities for image tag assignment, refinement, and tag-based image retrieval problems.

Argyrou, Giannoulaki & Tsapatsoulis (2018) applied topic modeling with latent dirichlet allocation on Instagram hashtags to predict the subject of the related images. Wu et al. (2018) designed an automatic image annotation model that creates semantically relevant, yet distinct and diverse tags.

Zhang, Hu, Hu & Yu (2020) initialized sparse coding based spatial pyramid matching and deep convolutional neural networks to model image features and metric learning techniques.

4. Image annotation based on the training dataset

Seeing the importance of training dataset in image annotation, methods are developed to improve it. Johnson, Ballan & Fei-Fei (2015) developed a model to improve multilabel image annotation. The model used image metadata nonparametrically to generate neighborhoods of related images using Jaccard similarities and deep neural network to blend visual information from the image and its neighbors. Wu, Yu, Huang & Yu (2015) modeled deep learning in a weakly supervised learning (multiple instance learning) frameworks and suggested that each image follows a dual multi-instance assumption. Feng & Bhanu (2015)

presented an approach to automatically generate intermediate image descriptors by exploiting concept co-occurrence patterns in the pre-labeled training set to depict complex scene images semantically.

Hu, Yang, Shen & Zhang (2017) designed a web image annotation system using multi-facet and structural knowledge. Srivastava & Srivastava (2018) reviewed the areas such as segmentation, feature extraction, template matching, and image classification. Wei, He & Zou (2020) presented content semantic image analysis and storage method based on intelligent computing of machine learning annotation.

5. Image annotation based on user interaction

Image annotation models are also generated based on user interaction. Pellegrin, Escalante, Gómez & González (2017) proposed various approaches for local and global image annotation intending to facilitate their organization and accessibility options. Sbaiz & Vetterli (2018) invented an application of image annotation for portable display devices, such as smartphones, to determine the positional placement of annotations.

Lin, Upchurch & Bala (2020) generated a model for semantic segmentation with sub-image decomposition to inpaint block-annotated images with high-quality labels without additional human efforts.

Motivation

The above literature review revealed that:

- *Nowadays, visual data are as common as textual data due to the explosive growth of digital technologies, ever-increasing visual data are created and stored. (Zhang, Islam & Lu, 2012).*

  The increasing need for the organizing of a huge set of images is the main motivation behind the research on image annotation. Due to the large amount of image data available worldwide, and increased storage capacity of the digital storage devices, there is an urgent need for an efficient annotation system to extract visual information on demand.

- *The reason to annotate data is to simplify access to it and allow for more effective searches. The problem with adding annotation manually is an extremely labor-intensive and time-consuming task. Automatic image annotation has thus emerged as an important research area due to its potential application on both image understanding and web image search (Hanbury,2006, Deljooi & Jassbi, 2015).*

  The main idea of automatic image annotation techniques is to automatically learn semantic concept models from a large number of image samples, and use these models to label new images. Thus there is a need to develop a hybrid image annotation system by using various intelligent methodologies for automatic image annotation.

- *A number of issues occur during the automatic image annotation process Wang & Li (2017); Verma & Jawahar(2017); Song et al. (2020) that affect the efficiency and effectiveness of decision making for the labeling of classes. Some challenges in image annotation are:*

✓ *Images with tiny objects like nails, handrails,*

✓ *Images with similar-looking objects like Apple and Ball, and*

✓ *Pixel Precision.*

An effort can be made to overcome these challenges. This proposed study is an effort in this direction. The distinguish feature of the proposed system will be the hybridization of the image annotation approaches as well as the hybridization of different machine learning algorithms.

This may overcome the challenges of existing automatic image annotation system.

PROPOSED WORK

The main goal of this research study is to develop a hybrid image annotation system for extracting information from the visual environment.

Objectives of the proposed study

- To prepare an image training dataset.
- To develop a tool for labeling multimodal data.
- To develop a hybrid methodology for automatic image annotation to provide a drawing interface.
- To incorporate the developed tools and methods in a decision support framework.
- To implement the developed methodology for a real-world situation.

Methodology To achieve the objectives

In the proposed study, an attempt is made to develop a hybrid image annotation system. The term hybridization refers to combining two or more approaches for designing the proposed system, for example including model-based approaches and learning approaches. The hybridization will be based on the collected image dataset. The major tasks of the proposed methodology are:

- In order to achieve 1st objective, an image dataset will be prepared by using the image features like type of image, resolution and image pre-processing steps will be performed on the dataset for cleaning, integrating, transforming and reducing the image data.
- To attain the 2nd objective, a labeling tool will be developed to label different categories of objects in an image using the python framework and computer vision modules along with PyQT for a graphical interface. This study will deal with the images based on the application area chosen for the research study.
- For attaining the 3rd objective, machine learning algorithms such as Regression, support vector machines, K-nearest neighbors will be used to develop a hybrid method for image annotation. This model will be utilized for the recognition of an object in an image.
- For achieving the 4th objective, the developed labeling tools and methods with in a decision support system framework to make it an intelligent image annotation tool. The features and algorithms for decision support system will be selected as per the study area.
- The system will be implemented for the education system. This system will help in improving the basic understanding of primary level students (all the boards and DEI) in any one subject say social science or biology.

Stages of building the proposed system

1. **Planning & requirement analysis**

   This study will plan a strategic analysis to evaluate the organization of image databases. Several designing, developing and refining iterations will take place according to business requirements. Requirements of domain experts will be identified to acquire a basic understanding of the application domain for formulating varied requirements. Then the business subject priorities will be established, depending on their relative importance, and feasibility of the necessary data. This priority list will be used to establish the subsequent iterations of the system development.

2. **Data collection**

   In this stage of the development cycle, digital images will be collected to form an image dataset. After preparing the image database, the image-related text will be extracted from primary/secondary data sources for obtaining the text to annotate the images. One of the possible ways may be to augment the image caption with image mention (i.e. discussion of the image).

3. **Training & Modeling**

   In this phase, training of the image datasets will be performed to provide the training and test set of images for the validation process.

4. **Design and development**

   Once the dataset is designed a conceptual framework of the proposed system will be developed. Labeling tools will be developed using techniques like machine learning, ontology, and/or multi-label decision making methods for recognizing the objects in the images. These tools will be incorporated in a decision support system for intelligent annotation. The system will be developed using python.

5. **Testing**

   This phase of the cycle will test the developed model and the labeling tool for image annotation. Efficiency analysis of the developed system will be done to compare the performance against the existing systems of automatic image annotation. Testing will be performed by taking the feedback and developing test cases.

   To conclude, the proposed system will

- Enable the primary level students to easily understand the concepts through image annotation systems which otherwise become difficult for them to learn.
- Be appreciable by the computer science researchers as it will improve the accuracy of predicting the images correctly.
- Provide additional enhanced experience for the decision-makers as machine learning-based trained models offer a completely changed and unified experience for the end-users.

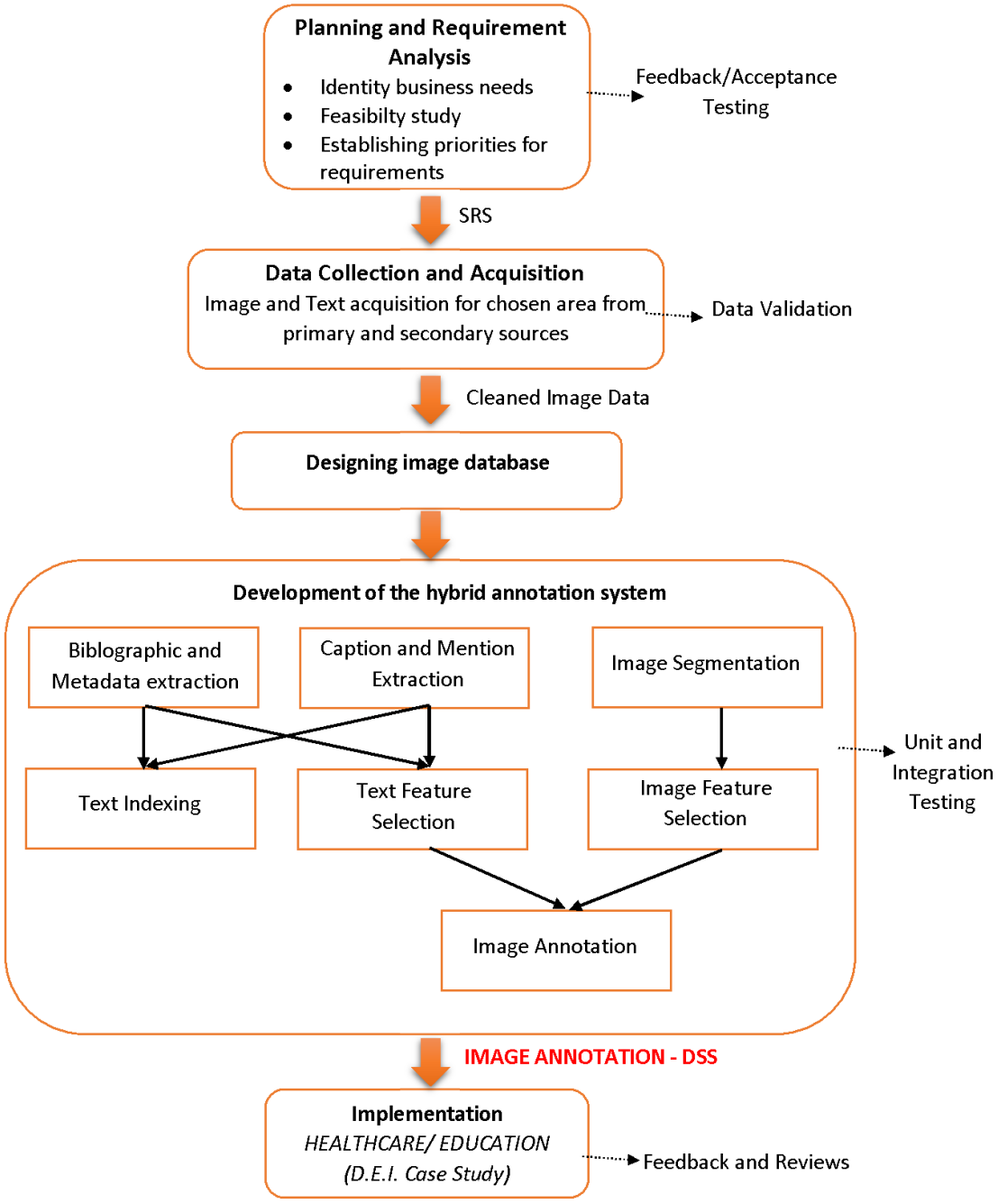

Figure 5: Stages of development

REFERENCES

Adikhresna, O., Kusumaningrum, R., & Warsito, B. (2019, October). Comparative Experimental Study of Multi-label Classification using Single Label Ground Truth with Application to Field Majoring Problem. In *2019 3rd Int . ernational Conference on Informatics and Computational Sciences (ICICoS)* (pp. 1-5). IEEE.

Albiero, V., KS, K., Vangara, K., Zhang, K., King, M. C., & Bowyer, K. W. (2020). Analysis of gender inequality in face recognition accuracy. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision Workshops* (pp. 81-89).

Ali, M. M., Hashim, N., & Hamid, A. S. A. (2020). Combination of laser-light backscattering imaging and computer vision for rapid determination of oil palm fresh fruit bunches maturity. *Computers and Electronics in Agriculture*, *169*, 105235.

Ali, M. M., Hashim, N., Bejo, S. K., & Shamsudin, R. (2020). Comparison of laser backscattering imaging and computer vision system for grading of seedless watermelons. *Journal of Food Measurement and Characterization*, *14*(1), 69-77.

Al-Osaim, F. R., & Bouguila, N. (2015, December). A finite gamma mixture model-based discriminative learning framework. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)* (pp. 819-824). IEEE.

Alqahtani, F. H., & Alsulaiman, F. A. (2020). Is image-based CAPTCHA secure against attacks based on machine learning? An experimental study. *Computers & Security*, *88*, 101635.

Amiri, S. H., & Jamzad, M. (2015). Automatic image annotation using semi-supervised generative modeling. *Pattern Recognition*, *48*(1), 174-188.

Argyrou, A., Giannoulakis, S., & Tsapatsoulis, N. (2018, September). Topic modeling on Instagram hashtags: An alternative way to automatic image annotation?. In *2018 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)* (pp. 61-67). IEEE.

Ballard, D. H. (1989, August). Reference Frames for Animate Vision. In *IJCAI* (Vol. 89, pp. 1635-1641).

Das, A., Degeling, M., Wang, X., Wang, J., Sadeh, N., & Satyanarayanan, M. (2017, July). Assisting users in a world full of cameras: A privacy-aware infrastructure for computer vision applications. In *2017*

IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (pp. 1387-1396). IEEE.

Deljooi, H., & Jassbi, S. J. (2015). A Multi-Criteria Decision Making Based Approach for Semantic Image Annotation. *International Journal of Computer-Aided Technologies*, *2*(1), 17-30.

Ding, X., Li, B., Xiong, W., Guo, W., Hu, W., & Wang, B. (2016). Multi-instance multi-label learning combining hierarchical context and its application to image annotation. *IEEE Transactions on Multimedia*, *18*(8), 1616-1627.

Fang, W., Ding, L., Love, P. E., Luo, H., Li, H., Pena-Mora, F., ... & Zhou, C. (2020). Computer vision applications in construction safety assurance. *Automation in Construction*, *110*, 103013.

Feng, L., & Bhanu, B. (2015). Semantic concept co-occurrence patterns for image annotation and retrieval. *IEEE transactions on pattern analysis and machine intelligence*, *38*(4), 785-799.

Feng, S., & Lang, C. (2018). Graph regularized low-rank feature mapping for multi-label learning with application to image annotation. *Multidimensional Systems and Signal Processing*, *29*(4), 1351-1372.

Guo, H. F., Han, L., Su, S., & Sun, Z. B. (2018). Deep multi-instance multi-label learning for image annotation. *International Journal of Pattern Recognition and Artificial Intelligence*, *32*(03), 1859005.

Hanbury, A. (2006). *Review of image annotation for the evaluation of computer vision algorithms*. Technical Report PRIP-TR-102, PRIP, TU Wien.

Hidajat, M. (2015). Annotation Based Image Retrieval using GMM and Spatial Related Object Approaches. *vol*, *8*, 399-408.

Hu, M., Yang, Y., Shen, F., Zhang, L., Shen, H. T., & Li, X. (2017). Robust web image annotation via exploring multi-facet and structural knowledge. *IEEE Transactions on Image Processing*, *26*(10), 4871-4884.

Huang, Q., Yang, L., Huang, H., Wu, T., & Lin, D. (2020). Caption-supervised face recognition: Training a state-of-the-art face model without manual annotation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (Vol. 1).

Im, D. H., & Park, G. D. (2015). Linked tag: image annotation using semantic relationships between image tags. *Multimedia Tools and Applications*, *74*(7), 2273-2287.

Ivasic-Kos, M., Pobar, M., & Ribaric, S. (2016). Two-tier image annotation model-based on a multi-label classifier and fuzzy-knowledge representation scheme. *Pattern recognition*, *52*, 287-305.

James, L. C., Marzloff, M. P., Barrett, N., Friedman, A., & Johnson, C. R. (2017). Changes in deep reef benthic community composition across a latitudinal and environmental gradient in temperate Eastern Australia. *Marine Ecology Progress Series*, *565*, 35-52.x

Ji, P., Gao, X., & Hu, X. (2017). Automatic image annotation by combining generative and discriminant models. *Neurocomputing*, *236*, 48-55.

Jin, C., & Jin, S. W. (2019). Multi-label automatic image annotation approach based on multiple improvement strategies. *IET Image Processing*, *13*(4), 623-633.

Jin, C., Liu, J., & Guo, J. (2015). A hybrid model-based on mutual information and support vector machine for automatic image annotation. In *Artificial intelligence perspectives and applications* (pp. 29-38). Springer, Cham.

Jin, J., & Nakayama, H. (2016, December). Annotation order matters: Recurrent image annotator for arbitrary length image tagging. In *2016 23rd International Conference on Pattern Recognition (ICPR)* (pp. 2452-2457). IEEE.

Jin, Y., Li, J., Ma, D., Guo, X., & Yu, H. (2017, July). A semi-automatic annotation technology for traffic scene image labeling based on deep learning preprocessing. In *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)* (Vol. 1, pp. 315-320). IEEE.

Jing, X. Y., Wu, F., Li, Z., Hu, R., & Zhang, D. (2016). Multi-label dictionary learning for image annotation. *IEEE Transactions on Image Processing*, *25*(6), 2712-2725.

Jiu, M., & Sahbi, H. (2017). Nonlinear deep kernel learning for image annotation. *IEEE Transactions on Image Processing*, *26*(4), 1820-1832.

Johnson, J., Ballan, L., & Fei-Fei, L. (2015). Love thy neighbors: Image annotation by exploiting image metadata. In *Proceedings of the IEEE international conference on computer vision* (pp. 4624-4632).

Ke, X., Zou, J., & Niu, Y. (2019). End-to-end automatic image annotation based on deep CNN and multi-label data augmentation. *IEEE Transactions on Multimedia*, *21*(8), 2093-2106.

Kohli, P., & Chadha, A. (2019, March). Enabling pedestrian safety using computer vision techniques: A case study of the 2018 uber inc. self-driving car crash. In *Future of Information and Communication Conference* (pp. 261-279). Springer, Cham.

Kromp, F., Bozsaky, E., Rifatbegovic, F., Fischer, L., Ambros, M., Berneder, M., ... & Beiske, K. (2020). An annotated fluorescence image dataset for training nuclear segmentation methods. *Scientific Data*, *7*(1), 1-8.

L. Pellegrin, H.J. Escalante, M. Montes-y Gómez, F.A. González, Local and global approaches for unsupervised image annotation, Multimedia Tools Appl. 76 (15) (2017) 16389–16414.

Li, K., Zou, C., Bu, S., Liang, Y., Zhang, J., & Gong, M. (2018). Multi-modal feature fusion for geographic image annotation. *Pattern Recognition*, *73*, 1-14.

Li, W., Yang, Y., Zhang, K., Long, E., He, L., Zhang, L., ... & Yun, D. (2020). Dense anatomical annotation of slit-lamp images improves the performance of deep learning for the diagnosis of ophthalmic disorders. *Nature Biomedical Engineering*, 1-11.

Li, X., Shen, B., Liu, B. D., & Zhang, Y. J. (2017). Ranking-preserving low-rank factorization for image annotation with missing labels. *IEEE Transactions on Multimedia*, *20*(5), 1169-1178.

Li, X., Uricchio, T., Ballan, L., Bertini, M., Snoek, C. G., & Bimbo, A. D. (2016). Socializing the semantic gap: A comparative survey on image tag assignment, refinement, and retrieval. *ACM Computing Surveys (CSUR)*, *49*(1), 1-39.

Lin, H., Upchurch, P., & Bala, K. (2020). Block Annotation: Better Image Annotation for Semantic Segmentation with Sub-Image Decomposition. *arXiv preprint arXiv:2002.06626*.

Liu, Y., Wen, K., Gao, Q., Gao, X., & Nie, F. (2018). SVM based multi-label learning with missing labels for image annotation. *Pattern Recognition*, *78*, 307-317.

Luo, Y., Liu, T., Tao, D., & Xu, C. (2015). Multiview matrix completion for multilabel image classification. *IEEE Transactions on Image Processing*, *24*(8), 2355-2368.

Ma, Y., Xie, Q., Liu, Y., & Xiong, S. (2019). A weighted KNN-based automatic image annotation method. *Neural Computing and Applications*, 1-12.

Maihami, V., & Yaghmaee, F. (2017). A review on the application of structured sparse representation at image annotation. *Artificial Intelligence Review*, *48*(3), 331-348.

Mehmood, Z., Mahmood, T., & Javid, M. A. (2018). Content-based image retrieval and semantic automatic image annotation based on the weighted average of triangular histograms using support vector machine. *Applied Intelligence*, *48*(1), 166-181.

Murthy, V. N., Maji, S., & Manmatha, R. (2015, June). Automatic image annotation using deep learning representations. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval* (pp. 603-606).

Pellegrin, L., Vanegas, J. A., Ovalle, J. E. A., Beltrán, V., Escalante, H. J., Montes-y-Gómez, M., & González, F. A. (2015). INAOE-UNAL at ImageCLEF 2015: Scalable Concept Image Annotation. In *CLEF (Working Notes)*.

Sagar, A., & Soundrapandiyan, R. (2020). Semantic Segmentation With Multi-Scale Spatial Attention For Self Driving Cars. *arXiv preprint arXiv:2007.12685*.

Sbaiz, L., & Vetterli, M. (2018). *U.S. Patent No. 9,953,438*. Washington, DC: U.S. Patent and Trademark Office.

Schleier, S., Stolz, N., Langner, H., Hasan, R., Roschke, C., & Ritter, M. (2020, July). Semi-automatic Annotation of OCT Images for CNN Training. In *International Conference on Human-Computer Interaction* (pp. 672-685). Springer, Cham.

Shapiro, L., & Stockman, G. (2001). Computer vision prentice hall. *Inc., New Jersey*.

Shi, Z., Yang, Y., Hospedales, T. M., & Xiang, T. (2016). Weakly-supervised image annotation and segmentation with objects and attributes. *IEEE transactions on pattern analysis and machine intelligence*, *39*(12), 2525-2538.

Song, H., Wang, P., Yun, J., Li, W., Xue, B., & Wu, G. (2020). A Weighted Topic Model Learned From Local Semantic Space for Automatic Image Annotation. *IEEE Access*, *8*, 76411-76422.

Song, J., Gao, L., Nie, F., Shen, H. T., Yan, Y., & Sebe, N. (2016). Optimized graph learning using partial tags and multiple features for image and video annotation. *IEEE Transactions on Image Processing*, *25*(11), 4999-5011.

Song, Q., Zhao, L., Luo, X., & Dou, X. (2017). Using deep learning for classification of lung nodules on computed tomography images. *Journal of healthcare engineering*, *2017*.

Srivastava, G., & Srivastava, R. (2018, January). A survey on automatic image captioning. In *International Conference on Mathematics and Computing* (pp. 74-83). Springer, Singapore.

Srivastava, G., & Srivastava, R. (2020). Design, Analysis, and Implementation of Efficient Framework for Image Annotation. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *16*(3), 1-24.

Sturman, O., von Ziegler, L., Schläppi, C., Akyol, F., Privitera, M., Slominski, D., & Bohacek, J. (2020). Deep learning-based behavioral analysis reaches human accuracy and is capable of outperforming commercial solutions. *Neuropsychopharmacology*, 1-11.

Tewari, A., Zollhofer, M., Kim, H., Garrido, P., Bernard, F., Perez, P., & Theobalt, C. (2017). Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 1274-1283).

Tian, D. (2018). Research on PLSA model-based semantic image analysis: a systematic review. *J. Inf. Hiding Multimed. Signal Process*, *9*(5), 1099-1113.

Trucco, E., & Verri, A. (1998). *Introductory techniques for 3-D computer vision* (Vol. 201). Englewood Cliffs: Prentice-Hall.

Uricchio, T., Ballan, L., Seidenari, L., & Del Bimbo, A. (2017). Automatic image annotation via label transfer in the semantic space. *Pattern Recognition*, *71*, 144-157.

Vatani, A., Ahvanooey, M. T., & Rahimi, M. (2020). An effective automatic image annotation model via attention model and data equilibrium. *arXiv preprint arXiv:2001.10590*.

Verma, Y., & Jawahar, C. V. (2017). Image annotation by propagating labels from semantic neighborhoods. *International Journal of Computer Vision*, *121*(1), 126-148.

Waheed, M., Akhtar, N., Afzal, H., Ahmed, M., Khan, D. M., Mufti, M. R., ... & Habib, A. (2020). Automatic Video Annotation of Human Health Care Action via Clustering. *Journal of Medical Imaging and Health Informatics*, *10*(10), 2512-2521.

Wang, J., & Li, G. (2017, June). A multi-modal hashing learning framework for automatic image annotation. In *2017 IEEE Second International Conference on Data Science in Cyberspace (DSC)* (pp. 14-21). IEEE.

Wang, J., Mueller, F., Bernard, F., & Theobalt, C. (2020). Generative Model-Based Loss to the Rescue: A Method to Overcome Annotation Errors for Depth-Based Hand Pose Estimation. *arXiv preprint arXiv:2007.03073*.

Wei, P., He, F., & Zou, Y. (2020). Content semantic image analysis and storage method based on intelligent computing of machine learning annotation. *Neural Computing and Applications*, 1-10.

Wu, B., Chen, W., Sun, P., Liu, W., Ghanem, B., & Lyu, S. (2018). Tagging like humans: Diverse and distinct image annotation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7967-7975).

Wu, B., Lyu, S., Hu, B. G., & Ji, Q. (2015). Multi-label learning with missing labels for image annotation and facial action unit recognition. *Pattern Recognition*, *48*(7), 2279-2289.

Wu, F., Wang, Z., Zhang, Z., Yang, Y., Luo, J., Zhu, W., & Zhuang, Y. (2015). Weakly semi-supervised deep learning for multi-label image annotation. *IEEE Transactions on Big Data*, *1*(3), 109-122.

Wu, J., Yu, Y., Huang, C., & Yu, K. (2015). Deep multiple instances learning for image classification and auto-annotation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3460-3469).

Xia, S., Chen, P., Zhang, J., Li, X., & Wang, B. (2017). Utilization of rotation-invariant uniform LBP histogram distribution and statistics of connected regions in automatic image annotation based on multi-label learning. *Neurocomputing*, *228*, 11-18.urc

Yakovlev, A., & Lisovychenko, O. (2020). An approach for image annotation automatization for artificial intelligence models learning. *Adaptive systems of automatic control*, *1*(36), 32-40.

Yin, S., & Bi, J. (2018, July). Medical image annotation based on deep transfer learning. In *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom), and IEEE Smart Data (SmartData)* (pp. 47-49). IEEE.

Zhang, D., Islam, M. M., & Lu, G. (2012). A review on automatic image annotation techniques. *Pattern Recognition*, *45*(1), 346-362.

Zhang, G., Hsu, C. H. R., Lai, H., & Zheng, X. (2018). Deep learning-based feature representation for automated skin histopathological image annotation. *Multimedia Tools and Applications*, *77*(8), 9849-9869.

Zhang, W., Hu, H., Hu, H., & Yu, J. (2020). Automatic image annotation via category labels. *Multimedia Tools and Applications*, 1-15.

Zhao, H. H., & Liu, H. (2020). Multiple classifiers fusion and CNN feature extraction for handwritten digits recognition. *Granular Computing*, *5*(3), 411-418.

Zheng, L., Caiming, Z., & Caixian, C. (2018). MMDF-LDA: An improved Multi-Modal Latent Dirichlet Allocation model for social image annotation. *Expert Systems with Applications*, *104*, 168-184.

Zurowietz, M., Langenkämper, D., Hosking, B., Ruhl, H. A., & Nattkemper, T. W. (2018). MAIA—A machine learning assisted image annotation method for environmental monitoring and exploration. *PloS one*, *13*(11), e0207498.